JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

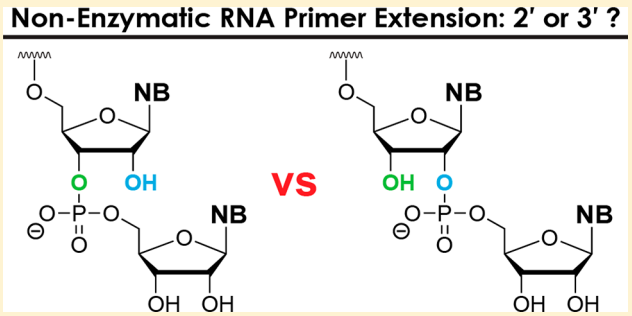
A Mechanistic Explanation for the Regioselectivity of Nonenzymatic RNA Primer Extension

Constantin Giurgiu, Li Li, Derek K. O'Flaherty, Chun Pong Tam, and Jack W. Szostak*

Howard Hughes Medical Institute, Department of Molecular Biology, and Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, United States

Supporting Information

ABSTRACT: A working model of nonenzymatic RNA primer extension could illuminate how prebiotic chemistry transitioned to biology. All currently known experimental reconstructions of nonenzymatic RNA primer extension yield a mixture of 2′-5′ and 3′-5′ internucleotide linkages. Although long seen as a major problem, the causes of the poor regioselectivity of the reaction are unknown. We used a combination of different leaving groups, nucleobases, and templating sequences to uncover the factors that yield selective formation of 3′-5′ internucleotide linkages. We found that fast and high yielding reactions selectively form 3′-5′ linkages. Additionally, in all cases with high 3′-5′ regioselectivity, Watson−Crick base pairing between the RNA monomers and the template is observed at the extension site and at the adjacent downstream position. Mismatched base-pairs and other factors that would perturb the geometry of the imidazolium bridged intermediate lower both the rate and regioselectivity of the reaction.



INTRODUCTION

Elucidating a pathway from prebiotic chemistry to Darwinian evolution is a fundamental problem in chemistry. Darwinian evolution requires a replicator that can propagate genetic information,[1] and RNA is an promising candidate for such a task. RNA stores information in its sequence of four bases, and folds into complex three-dimensional structures capable of catalysis and molecular recognition. In 1968 Crick, Orgel and Woese proposed that all life on Earth started from RNA, a scenario that later became known as the "RNA World".[2−4] In this scenario, early life forms had an RNA genome that encoded RNA enzymes.[5] The original RNA world hypothesis suggested that genomic replication would be a RNA catalyzed process, executed by a polymerase ribozyme. Despite recent progress,[6,7] major improvements in the accuracy and efficiency of such ribozymes are needed to permit self-replication. A plausible mechanism through which a RNA-based system could have evolved efficient polymerase ribozymes is through cycles of nonenzymatic RNA replication.[8] Over the last 50 years, considerable effort has gone into achieving nonenzymatic RNA replication in the laboratory, with little success. These attempts have highlighted several major issues that must be understood to enable genetic replication in an RNA world setting.[8] One such problem is the poor regioselectivity of the template directed-synthesis of RNA: the phosphodiester bond connecting the ribonucleotides can form between the 5′ phosphate group of one nucleotide and the hydroxyl group on either the 2′ or 3′ carbon atom of the next nucleotide. In biology, templated RNA synthesis is performed by polymerase enzymes, which exclusively form 3′-5′ phosphodiester bonds. In the absence of strict steric constraints imposed by enzyme active sites, a mixture of products is obtained.[9,10]

The formation of 2′-5′ linkages was long considered a fatal flaw of template-directed RNA synthesis because it was thought that linkage heterogeneity would impede the evolution of functional RNA molecules.[8] However, our group has shown that nucleic acids with nonheritable backbone heterogeneity can generate nucleotide binding aptamers through in vitro evolution.[11] Furthermore, some RNA aptamers and ribozymes containing up to 25% 2′-5′ linkages retain molecular recognition and catalytic properties.[12] These recent findings suggest that the poor regioselectivity of nonenzymatic RNA primer extension is not an insurmountable hurdle to chemical RNA replication. On the contrary, backbone heterogeneity assists with another difficult challenge, the strand separation problem. RNA duplexes of over 30 base pairs are unlikely to thermally denature in conditions compatible with the copying chemistry. Therefore, a primer-template complex that is efficiently extended inhibits further rounds of template-directed synthesis. Backbone heterogeneity lowers the melting point of RNA duplexes, allowing strands to separate at lower temperatures.[12,13] In addition, 2′-5′ linked RNA strands retain templating abilities. Switzer and colleagues have shown that primer extension proceeds in high yield, but with a lower rate, on a 2′-5′ linked template.[14] As a result, backbone

Received: August 17, 2017
Published: November 7, 2017

DOI: 10.1021/jacs.7b08784
J. Am. Chem. Soc. 2017, 139, 16741−16747

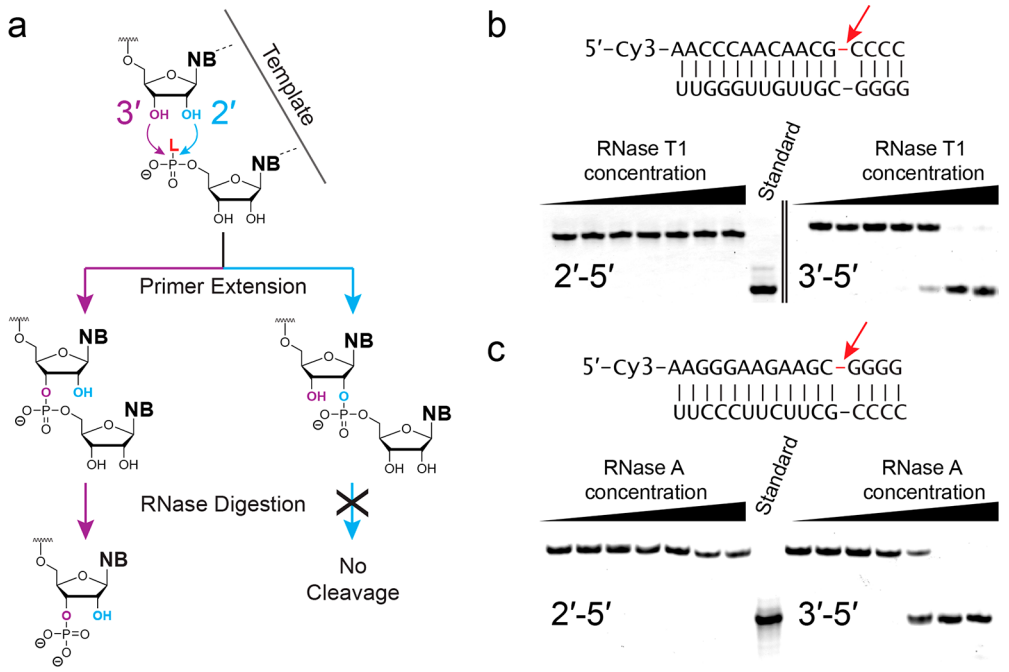


Figure 1. An assay for determining the percentage of 2′-5′ linkages in an oligonucleotide. (a) Templated RNA primer extension forms either a 2′-5′ linkage (cyan pathway) or a 3′-5′ linkage (magenta pathway). Endoribonucleases selectively cleave 3′-5′ linkages; NB stands for nucleobase; L represents an imidazole leaving group. (b) Top: sequence of the primer-template duplex used in the RNase T1 assay—if the red bond indicated by the arrow is a 3′-5′ linkage it will be cleaved by the enzyme. Bottom: PAGE analysis of the enzymatic digest confirms that a 2′-5′ extended primer is not cleaved by the enzyme at any of the concentrations tested, while the 3′-5′ linkage is fully cleaved at high RNase concentration. Standard refers to the synthetically obtained 3′-phosphate primer. Assay dynamic range and detailed experimental conditions can be found in the Supporting Information. (c) Top: sequence of the primer-template duplex used in the RNase A assay, notations identical to panel (b); Bottom: PAGE analysis of the RNase A assay, analogous to panel (b) bottom.

heterogeneity potentially assisted the evolution of function in the early stages of the RNA world.

The regioselectivity of the reaction is affected by the metal catalyst,[15] the leaving group on the activated monomers[16] and by using activated oligomers instead of monomers.[17] Despite extensive efforts, it is not known precisely how these factors operate.

Our recent discovery of a superior leaving group for RNA primer extension[18] and of an alternative mechanism for the reaction[19] has prompted us to revisit the regioselectivity problem. In this mechanism, two monomers react with each other to form an imidazolium-bridged dimer, which then presumably binds to the template at two adjacent sites. Here, we examined the influence of the leaving group, nucleobase and of the templating sequence on the regioselectivity of chemical RNA primer extension. Surprisingly, fast and efficient reactions preferentially formed 3′-5′ linkages. In all instances where we observed high 3′-5′ regioselectivity the following two criteria were invariably met: (1) Watson−Crick base pairing was observed at the extension site and at the adjacent upstream and downstream positions, and (2) the downstream binding nucleotides had a leaving group on their 5′ end. Unnatural 2′-5′ linkages are formed in RNA primer extension during the copying of the last base of the template, through mismatched copying, and when multiple adenosine or uridine monomers are added sequentially. The relatively low frequency of such events potentially leads to a proportion of 2′-5′ linkages that lowers the melting temperature of the RNA duplex, without preventing the evolution of functional RNA sequences. This leads us to consider that the backbone heterogeneity arising from chemical RNA primer extension was not an issue in the origin of life.

RESULTS

We have adapted Orgel's well established method of quantifying the regioselectivity of templated nonenzymatic RNA primer extension.[20] In this method, the RNA primer was extended by the templated addition of 5′-phosphorimidazolide monomers (Figure 1a). The reaction products were then treated with either RNase T1 (Figure 1b) or RNase A (Figure 1c), endoribonucleases that specifically cleave 3′-5′ linkages to 3′-phosphates after guanosine residues, and pyrimidine residues, respectively. The sequences of the primers were designed such that the enzymes cleave at a single site, after the first bond that is formed during the primer extension reaction. Thus, if the first added residue is connected through a 3′-5′ linkage, the primer would be cleaved. In contrast, if the first residue is connected through a 2′-5′ linkage the extension products would be insensitive to cleavage. The RNase T1 digestion assay can detect as little as 6%, while the RNase A assay can detect as little as 2% 2′-5′ linkages (Figures S1 and S2).

We first examined how the nucleobase on the incoming monomer affects the ratio of 2′-5′ to 3′-5′ linkages (Figure 2). Our group has previously identified 2-aminoimidazole activated 5′-nucleotide monophosphates (2-AImpN) as the most efficient monomers for RNA polymerization to date.[18] Primer extension with 2-AImpG **2a** and 2-AImpC **3a** monomers was fast and regioselective, and the proportion of 2′-5′ linkages formed was below the detection limit of the assays. In addition, the regioselectivity of the reaction was identical over a wide

DOI: 10.1021/jacs.7b08784
J. Am. Chem. Soc. 2017, 139, 16741−16747

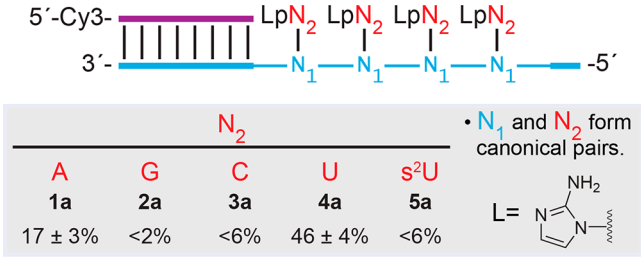


Figure 2. Nucleobase influence on regioselectivity. Enzymatic digestion was carried out with RNase A for the purine monomers and RNase T1 for the pyrimidine monomers. The percentages represent the proportion of 2′-5′ linkages formed and were obtained from four independent experiments.

range of concentrations of the activated monomer **2a** (Figure S3).

However, 2-AImpA **1a** and 2-AImpU **4a** additions were considerably slower and formed 17% and 46% 2′-5′ linkages, respectively. A and U pair through two hydrogen bonds while G and C pair through three hydrogen bonds, suggesting that the strength of the Watson−Crick interaction plays an important role in the regioselectivity of the reaction. We next measured the regioselectivity of the primer extension using 2-thiouridine (s^2U) phosphorimidazolide monomer **5a**. If base pair stability is important, the greater thermodynamic stability of the s^2U:A base pair compared to the canonical A:U pair should decrease the proportion of 2′-5′ linkages. The s^2U activated monomer **5a** formed no detectable 2′-5′ linkages, supporting this hypothesis.

The regioselectivity of chemical RNA primer extension depends on the imidazole leaving group used.[21,22] For example, reactions using imidazole activated monomers are biased toward the formation of 2′-5′ linkages,[23] whereas the 2-methylimidazole group is selective for 3′-5′ linkages.[22] We examined how the leaving group influences the regioselectivity of the reaction when using different guanosine monomers. The proportion of 2′-5′ linkages obtained correlates with the rate of extension; the more reactive 2-aminoimidazole and 2-methylimidazole leaving groups led to 3′-5′ linkages exclusively, while the slowest of the series, imidazole, produced 6% 2′-5′ linkages (Figure 3).

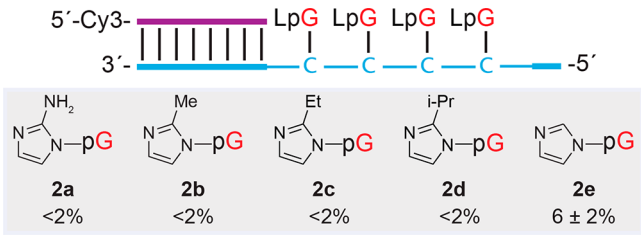


Figure 3. Leaving group influence on regioselectivity. Enzymatic digestion was carried out with RNase A. The percentages represent the proportion of 2′-5′ linkages formed and were obtained from four independent experiments.

The nature of the leaving group does not have as large an influence on the regioselectivity as on the rate of the reaction. Although there is a 7-fold difference in reaction rates between **2a** and **2e**, both leaving groups form less than 10% 2′-5′ linkages. The identity of the nucleobase seems to be the main determinant of regioselectivity. Thus, the recognition of the monomers by the template is crucial for a regioselective reaction. However, the Orgel group has shown that the Pb^{2+} catalyzed polycytidilic acid templated oligomerization of guanosine monomer **2e** produces 2′-5′ linked polyguanylic acid with remarkably fast rates.[24,25] The difference in regioselectivity observed between our experiments and the Orgel experiments presumably arises from the use of a different metal catalyst and possibly the formation of RNA triplex structures in the latter.

Assuming the template and monomer interact in a canonical Watson−Crick fashion, we investigated the cases in which such a pairing was not possible. Wu and Orgel reported that primer extension with 2-methylimidazole activated 5′-guanosine phosphorimidazolide **2b** on a G template results in substantial 2′-5′ linkage formation.[26] We explored the different possibilities of mismatch incorporation, using 2-aminoimidazole activated nucleotide monophosphates (Figure 4).

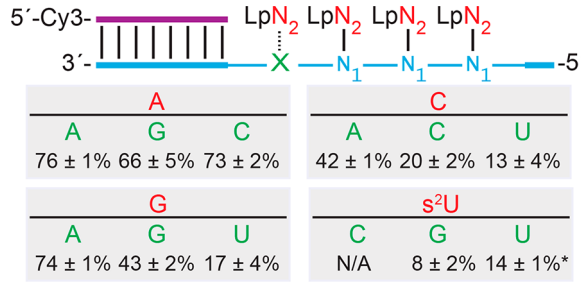


Figure 4. Mismatches lead to decreased regioselectivity. (a) RNase A was used for the enzymatic digestion of purine monomers, and RNase T1 was used for pyrimidine monomers. The nucleotides N_1 and N_2 form canonical base pairs; X and N_2 form mispairs. L stands for the 2-aminoimidazole leaving group. The percentages represent the proportion of 2′-5′ linkages formed and were obtained from four independent experiments. No extension products were observed when the s^2U monomer was added across a templating cytosine residue. *The value was obtained from three independent experiments.

The regioselectivity of the reaction decreased when compared to the case of canonical base pair formation. Purine-purine mismatches form the highest percentages of 2′-5′ linkages we have observed in any condition tested. Pyrimidine-pyrimidine mismatches form approximately 3-fold less 2′-5′ linkages. A possible explanation for this phenomenon is that the template can form a one nucleotide loop when the first templating nucleotide is a pyrimidine, and the products obtained are the result of frameshift mutations. Interestingly, the 2-thiouridine monomer **5a** adds to the primer in poor yields in all three cases examined (no addition vas observed when the templating nucleotide was cytosine), but with high 3′-5′ regioselectivity.

The rate of nonenzymatic primer extension decreases in the absence of activated downstream binding nucleotides.[26] To investigate the effect of downstream binding on regioselectivity, we designed a primer-template combination in which a single nucleotide is available to template the extension (Figure 5). In this case all three nucleotides (A, G, C) formed between 15 to 49% of 2′-5′ linkages. Similar to the experiments presented in Figure 2 the regioselectivity did not differ greatly between the 2-methylimidazole and 2-aminoimidazole leaving groups, for any given nucleobase. However, for the imidazole derivative **2e**, twice the fraction of 2′-5′ linkages formed when compared to the faster and more efficient **2a** and **2b** monomers.

DOI: 10.1021/jacs.7b08784
J. Am. Chem. Soc. 2017, 139, 16741−16747

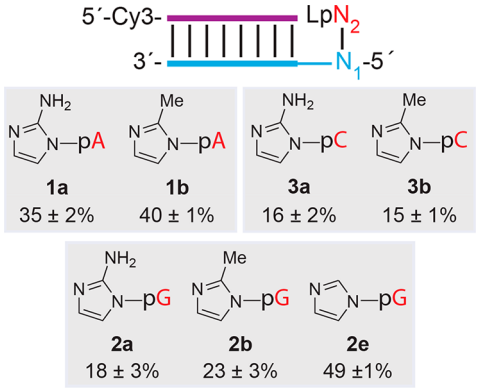


Figure 5. Copying of the last nucleotide generates backbone heterogeneity. For G and A monomers enzymatic digestion was carried out with RNase A; for the C monomers RNase T1 was used. The percentages represent the proportion of 2′-5′ linkages formed and were obtained from four independent experiments.

Copying of the last nucleotide of a template also resulted in decreased regioselectivity, which implies that downstream binding nucleotides promote 3′-5′ regioselective reactions. The mechanism through which downstream binders influence the reaction could involve base stacking, leaving group−leaving group interactions, or a mixture of both. Alternatively, two activated monomers can interact covalently to form an imidazole-bridged dimer,[19] and the conformational constraint resulting from the formation of two base pairs could influence regioselectivity. To better understand how this interaction influences the regioselectivity of the reactions, we varied the leaving groups on the adenosine monomers and on the downstream binding guanosine helpers when copying a UCCC template (Figure 6a). Unexpectedly, almost all of the combinations of imidazole activating groups we have tested formed 3′-5′ linkages exclusively. When the adenosine monomer was omitted, the rate, yield and regioselectivity of the reaction decreased dramatically, suggesting that the correct monomer is added (Figure S4). Additionally, when the reaction was carried out with a mixture of all four canonical nucleotides, all of the products were 3′-5′ linked. Considering that the mismatch addition opposite a templating uracil residue formed at least 13% of 2′-5′ linkages, and that the incorporation of a mismatched nucleotide causes a sharp decrease in rate,[27] we conclude that the fidelity of the reaction was high.

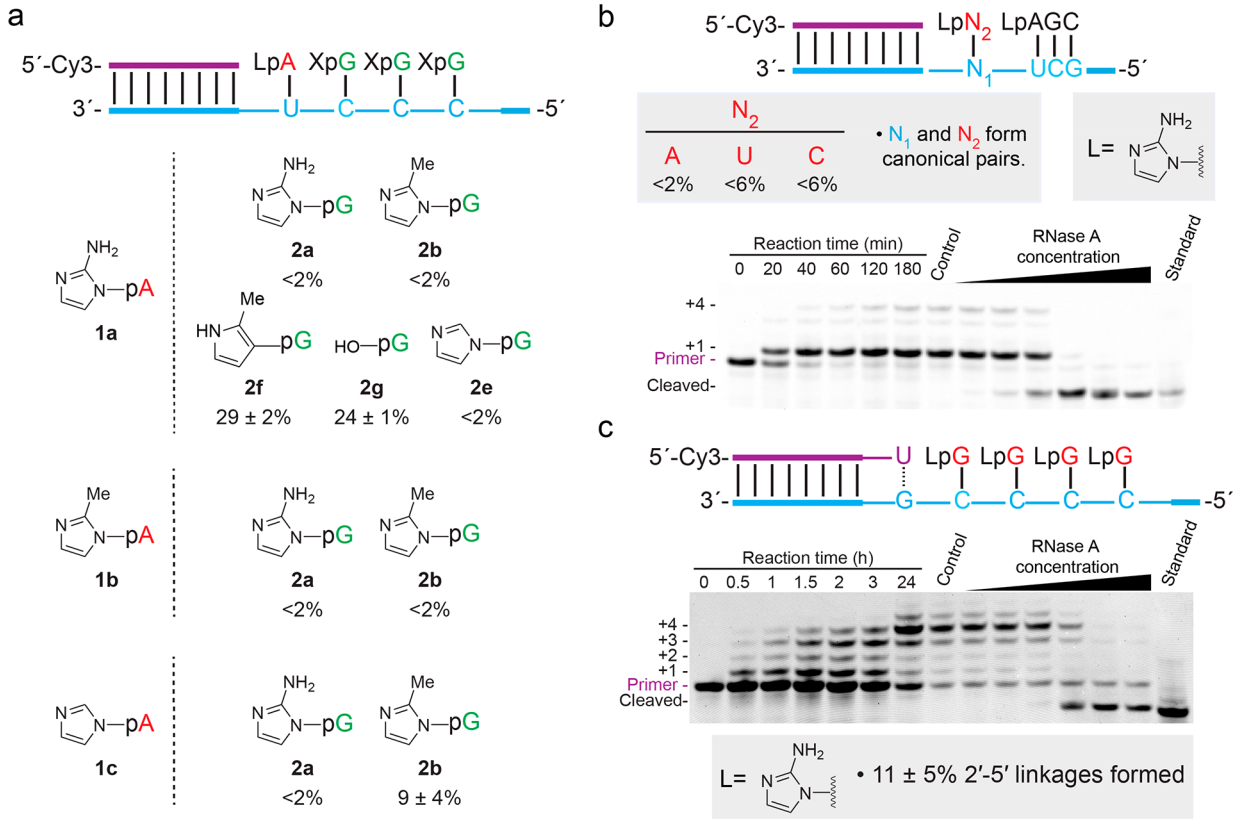


Figure 6. (a) Downstream binders improve the regioselectivity. In all cases RNase A was used for enzymatic digestion. The percentages represent the proportion of 2′-5′ linkages formed in the reactions using the A monomer on the left in combination with the G monomer above the numeric value. The values were obtained from four independent experiments. (b) Trimer-assisted primer extension. RNase T1 was used for enzymatic digestion of the U and C monomers, while RNase A was used with the A monomer. The percentages represent the proportion of 2′-5′ linkages formed and were obtained from four independent experiments. The electrophoretogram of the primer extension reaction and subsequent enzymatic digestion when 2-AImpU was used together with 2-AImpAGC. Control refers to the omission of the enzyme; Standard refers to the 3′-phosphate primer. (c) Primer template wobble pair moderately affects regioselectivity. The dashed line in the primer-template duplex represents the wobble-pair formed between G and U. The electrophoretogram shows the results of the primer extension and RNase A enzymatic digestion. Control refers to omission of the enzyme; standard refers to the 3′-phosphate primer. The percentage value represents the proportion of 2′-5′ linkages formed and was obtained from four independent experiments.

DOI: 10.1021/jacs.7b08784
J. Am. Chem. Soc. 2017, 139, 16741−16747

The regioselectivity suffered when no leaving group was present on the downstream binding nucleotides **2g**, or when a nonhydrolyzable analogue of 2-methylimidazole **2f** was used.

Leaving groups on neighboring nucleotides could interact via π-stacking, through the formation of hydrogen bonds between a protonated imidazole group and an adjacent unprotonated one, or through a cation-π interaction between two such groups (Figure S5). Although the 2-methylimidazole analogue **2f** can participate in these interactions, it forms 29% of 2′-5′ linkages when used as a helper. In fact, there is little difference in the proportion of 2′-5′ linkages formed between the analogue **2f** and guanosine monophosphate **2g**. Thus, hydrogen bonding and stacking interactions between leaving groups do not fully determine the regioselectivity of the reaction.

Downstream binding 5′-activated oligonucleotides can enhance the rate and fidelity of primer extension,[28] especially in troublesome A and U rich regions. Having determined that activated downstream monomers induce the formation of 3′-5′ linkages we tested whether an activated oligonucleotide would similarly affect the regioselectivity of the reaction. In the presence of the 5′-phosphorimidazolide trimers, both A and U monomers added without forming any 2′-5′ linkages (Figure 6b). Furthermore, the highly regioselective addition of 2-AImpC is unaffected by the presence of the trimer. Thus, downstream binders not only enhance the rate of the copying reaction for A and U rich regions, but also favor the formation of 3′-5′ internucleotide linkages.

We next investigated the effect of a fraying base pair at the 3′ end of the primer. This situation appears in RNA primer extension when a mismatched nucleotide is added to the primer. The rates of reaction postmismatch are an order of magnitude slower,[27] which lead us to hypothesize that the regioselectivity of the reaction should decrease concomitantly. Introducing a C/A mismatch at the 3′-end of the primer significantly reduced the regioselectivity of 2AImpG **2a** addition; 2′-5′ linkages were formed in 40% yield (Figure S6). Replacing the C/A mismatch with a G/U wobble pair improves the regioselectivity of the reaction. The reaction was slower than in the case of a G/C canonical pair, but the regioselectivity was similar (Figure 6c). Only 11% of 2′-5′ linkages are obtained, 4-times less than in the case of the A/C mismatch.

The formation of 2′-5′ linkages in the primer extension reaction raises the question of their heritability over multiple rounds of primer extension.

Prakash and Switzer have shown that primer extension on a template that has four consecutive 2′-5′ linkages immediately after the primer annealing site proceeds 17 times slower than on the fully 3′-5′ linked template. The reaction produces an equal mixture of 2′-5′ and 3′-5′ linkages.[14] We examined the effect of a single 2′-5′ linkage immediately after the primer annealing site (Figure 7). Although the reaction is 5.5 times slower than with the 3′-5′ linked template, the regioselectivity is high: no detectable 2′-5′ linkages were formed. Therefore, a single 2′-5′ linkage in the template does not get passed on to the daughter strands, and is lost after a round of chemical RNA primer extension.

DISCUSSION

The mechanism of template directed primer extension with 5′-phosphorimidazolide substrates was assumed to involve the binding of the activated monomer to the template, followed by a nucleophilic attack by either the 2′ or 3′ ribose hydroxyl on the phosphate group of the monomer. Our group recently suggested that an alternative mechanism is potentially operating.[19] In this mechanism, two monomers react with each other to form an imidazolium-bridged dimer, which then binds to the template, presumably at two adjacent sites (Figure 8). Our experiments show that Watson−Crick base pairing at both the reaction site and the site immediately downstream is required for 3′-5′ regioselectivity, a result consistent with the binding of an imidazolium-bridged dinucleotide intermediate. Therefore, the poor regioselectivity observed with A:U base pairs was surprising, since a dinucleotide intermediate binding to the template by two Watson−Crick base pairs would be expected to give the same constrained geometry as a C−C or G−G intermediate, thus high regioselectivity. Possibly, the weak stacking interactions between two adjacent U nucleotides in either the imidazolium bridged intermediate or the template lead to a disordered geometry of the reaction center. Conversely, the high regioselectivity of the 2-thiouridine monomer (**5a**) can be explained by its enhanced preorganization in the 3′-endo conformation,[29] which leads to a more rigid geometry of the reaction center. The poor regioselectivity and reactivity of the A and U monomers could stem from a common cause: poor conformational constraint of the reaction center. Indeed, we were able to observe a strong correlation between the yield of a reaction and its preference for the formation of 3′-5′ linkages (see Figure S7 for a detailed discussion). This explanation is also consistent with the loss of

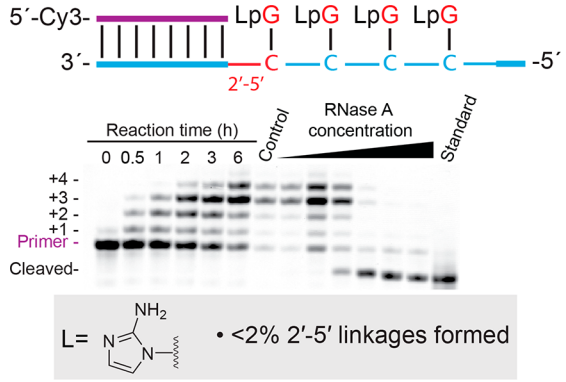


Figure 7. A 2′-5′ linkage in the template is not heritable. The electrophoretogram shows the results of the primer extension RNase A enzymatic digestion. Control refers to omission of the enzyme; standard refers to the 3′-phosphate primer. The percentage value represents the proportion of 2′-5′ linkages formed and was obtained from four independent replicates.

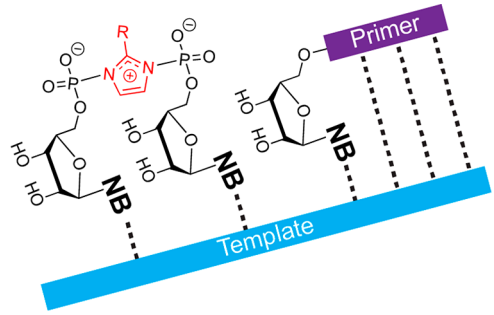


Figure 8. Imidazolium bridged dimer presumably binds to the template through Watson−Crick base pairs at two adjacent sites. The imidazolium core is highlighted in red.

DOI: 10.1021/jacs.7b08784
J. Am. Chem. Soc. 2017, 139, 16741−16747

reactivity and selectivity observed with other perturbations such as mismatches. Although not sufficient for a regioselective reaction, the formation and binding of an imidazolium bridged dinucleotide intermediate preorganizes the reaction center, and increases the proportion of 3′-5′ linkages formed. For example, in cases where the formation of a heterodimer intermediate is required (Figure 6a), we see a marked increase in the proportion of 2′-5′ linkages formed when one of the monomers cannot participate in the dimer formation reaction. Thus, when 2-AImpA is used together with guanosine monophosphate **2g**, or the nonhydrolyzable analogue **2f**, only the AA bridged dimer can be formed, but not the AG dimer that would correctly pair with the template. Consequently, the regioselectivity of the reaction decreases considerably compared to the cases in which the AG dimer is easily formed.

The nature of the internucleotide linkage formed in nonenzymatic primer extension is determined in part by the relative reactivity of the ribose 2′ and 3′ hydroxyl groups. Two independent studies looked at the distribution of 3′-5′ and 2′-5′ internucleotide linkages when 5′-phosphorimidazolide monomers reacted with phosphate-capped mononucleotides, in the absence of a template.[30,31] In these studies, the 2′-hydroxyl group was more reactive than its 3′-counterpart across all nucleotide and imidazole leaving group combinations. We find that the reactivity of the hydroxyl groups is reversed when the nucleophile is contained in a RNA duplex. Presumably in the primer-template duplex the accessibility toward electrophiles of the 3′-OH group is increased at the expense of the 2′-OH group. Additionally, our current findings show that the leaving group of the activated monomers has little influence on the regioselectivity of the primer extension. This result agrees with a previous study from our group,[17] which shows that the RNA-templated ligation of activated oligonucleotides displays a strong preference for 3′-5′ linkages, even when the leaving group was inorganic pyrophosphate. Presumably, the high regioselectivity of templated ligation and primer extension reactions is a consequence of the reactions occurring in a preorganized extended Watson−Crick duplex.

Prakash and Switzer showed that primer extension on a fully 2′-5′ linked template forms an equal mixture of 2′-5′ and 3′-5′ linkages.[14] Here we show that fast reactions on a 3′-5′ linked template proceed with high 3′-5′ regioselectivity. Additionally, we show that a single 2′-5′ linkage in the template does not affect the high 3′-5′ regioselectivity of the reaction. These two results suggest that a 2′-5′ linkage is not heritable, while for efficient reactions, 3′-5′ linkages will carry over to the daughter strands. Furthermore, since the extension across a 2′-5′ linkage is 5-fold slower than for 3′-5′ linkage, the fully 3′-5′ templates will outcompete the linkage heterogeneous ones. These observations are important to the origin of life, because all known processes through which RNA monomers polymerize in the absence of a template form a considerable proportion of 2′-5′ linkages.[32−34] Consequently, after multiple rounds of nonenzymatic RNA primer extension, the 3′-5′ linkages will be enriched whereas the 2′-5′ linkages will be depleted.

The Sutherland group recently demonstrated a mechanism for enriching 3′-5′ linked RNA starting from a pool of RNAs containing mixed linkages.[35] The 2′-5′ linkage is more hydrolytically labile than the 3′-5′ linkage.[36] Acetylation of a mixture of 2′ and 3′ terminal RNA phosphates is selective for the 2′ hydroxyl groups and subsequent templated ligation then forms 3′-5′ linked RNA.[37] If the hydrolysis and ligation processes are coupled in an energy dissipative cycling process,[38] they yield a plausible mechanism for the enrichment of 3′-5′ linkages. We show here that simple copying chemistry inherently favors 3′-5′ linkages. However, considering the difference in hydrolytic stability between the 3′-5′ and 2′-5′ phosphodiester linkages, a RNA duplex containing multiple 2′-5′ linkages in each strand would be hydrolyzed during the repeated heating−cooling cycles presumably required to enable multiple rounds of primer extension. The recycling chemistry of Sutherland et al. could enable the repair of such hydrolytic damage, while simple copying chemistry would be unable to do so. There is also the possibility that hydrolysis is slow, for example at low Mg^{2+} levels. Additionally, if strand displacement synthesis is possible, the need for temperature cycling is alleviated. Under such conditions, our results suggest that cycles of replication would lead to the depletion of 2′-5′ linkages. Of course, the two approaches are not mutually exclusive and may both have operated to some extent.

Backbone heterogeneity will be generated in the event of a mismatch or when a terminal nucleotide is copied. The ligation of such oligonucleotides will form long RNA strands with interspersed 2′-5′ linkages, which could evolve function and have lower duplex melting temperature than the canonically linked isomers. Thus, the balance of the two opposing selection forces i. e. fast replication and ease of strand displacement, will determine the percentage of 2′-5′ linkages in a pool of RNA molecules.

CONCLUSION

We observe that the rate and yield of chemical RNA primer extension are correlated with its regioselectivity. The reactions that rapidly proceed to completion contain mostly 3′-5′ linkages. Both the high reactivity and the regioselectivity of the reaction have a common origin: a properly preorganized geometry in the reaction center. In all instances of fast and regioselective reactions the RNA monomers can form Watson−Crick base pairs with the template at the extension site and the adjacent downstream position. In addition, the downstream binding nucleotide must have a leaving group on its 5′ phosphate. Our results are consistent with a reaction mechanism involving a 5′-5′ imidazolium bridged intermediate formed between the incoming monomer and a downstream monomer or an oligonucleotide. Reaction conditions that would perturb the binding of the intermediate to the template lead to a simultaneous decrease in reaction rate and regioselectivity.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b08784.

Methods and materials; Figures S1−S7; supplementary discussion; Table S1 (PDF)

AUTHOR INFORMATION

Corresponding Author

*szostak@molbio.mgh.harvard.edu

ORCID

Constantin Giurgiu: 0000-0003-0145-0110
Li Li: 0000-0003-4766-5782
Chun Pong Tam: 0000-0001-6381-9011
Jack W. Szostak: 0000-0003-4131-1203

DOI: 10.1021/jacs.7b08784
J. Am. Chem. Soc. 2017, 139, 16741−16747

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

J.W.S is an investigator of the Howard Hughes Medical Institute. L.L. is a Life Sciences Research Foundation Fellow. D.K.O. is a recipient of a Postdoctoral Research Scholarship from the Fonds de recherche du Québec—Nature et technologies (FRQNT), Quebec, Canada. This work was supported in part by grants from the Simons Foundation to J.W.S. (290363) and from the NSF (CHE-1607034) to J.W.S. The authors would like to thank Z. Todd and D. Duzdevich for helpful discussions.

■ REFERENCES

(1) Szostak, J.; Bartel, D.; Luisi, P. *Nature* **2001**, *409*, 387−390.
(2) Crick, F. H. C. *J. Mol. Biol.* **1968**, *38*, 367−379.
(3) Orgel, L. E. *J. Mol. Biol.* **1968**, *38*, 381−393.
(4) Woese, C. *The Genetic Code*; Harper and Row: New York, 1968.
(5) Gilbert, W. *Nature* **1986**, *319*, 618.
(6) Horning, D. P.; Joyce, G. F. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 9786−9791.
(7) Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P. *Science* **2011**, *332*, 209−212.
(8) Szostak, J. W. *J. Syst. Chem.* **2012**, *3*, 2.
(9) Kozlov, I. A.; Orgel, L. E. *Mol. Biol.* **2000**, *34*, 781−789.
(10) Joyce, G. F. *Cold Spring Harbor Symp. Quant. Biol.* **1987**, *52*, 41−51.
(11) Trevino, S. G.; Zhang, N.; Elenko, M. P.; Lupták, A.; Szostak, J. W. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 13492−13497.
(12) Engelhart, A. E.; Powner, M. W.; Szostak, J. W. *Nat. Chem.* **2013**, *5*, 390−394.
(13) Giannaris, P. A.; Damha, M. J. *Nucleic Acids Res.* **1993**, *21*, 4742−4749.
(14) Prakash, T. P.; Roberts, C.; Switzer, C. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 1522−1523.
(15) Bridson, P. K.; Orgel, L. E. *J. Mol. Biol.* **1980**, *144*, 567−577.
(16) Joyce, G. F.; Inoue, T.; Orgel, L. E. *J. Mol. Biol.* **1984**, *176*, 279−306.
(17) Rohatgi, R.; Bartel, D. P.; Szostak, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 3340−3344.
(18) Li, L.; Prywes, N.; Tam, C. P.; O'Flaherty, D. K.; Lelyveld, V. S.; Izgu, E. C.; Pal, A.; Szostak, J. W. *J. Am. Chem. Soc.* **2017**, *139*, 1810−1813.
(19) Walton, T.; Szostak, J. W. *J. Am. Chem. Soc.* **2016**, *138*, 11996−12002.
(20) Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992**, *114*, 317−322.
(21) Inoue, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1981**, *103*, 7666−7667.
(22) Inoue, T.; Orgel, L. E. *J. Mol. Biol.* **1982**, *162*, 201−217.
(23) Weimann, B. J.; Lohrmann, R.; Orgel, L. E.; Schneider-Bernloehr, H.; Sulston, J. E. *Science* **1968**, *161*, 387.
(24) Lohrmann, R.; Orgel, L. E. *J. Mol. Biol.* **1980**, *142*, 555−567.
(25) Lohrmann, R.; Bridson, P. K.; Orgel, L. E. *Science* **1980**, *208*, 1464−1465.
(26) Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992**, *114*, 5496−5501.
(27) Leu, K.; Kervio, E.; Obermayer, B.; Turk-MacLeod, R. M.; Yuan, C.; Luevano, J.-M.; Chen, E.; Gerland, U.; Richert, C.; Chen, I. A. *J. Am. Chem. Soc.* **2013**, *135*, 354−366.
(28) Prywes, N.; Blain, J. C.; Del Frate, F.; Szostak, J. W. *eLife* **2016**, e17756 DOI: 10.7554/eLife.17756.
(29) Agris, P. F.; Sierzputowska-Gracz, H.; Smith, W.; Malkiewicz, A.; Sochacka, E.; Nawrot, B.; Malkiewicqt, A.; Nawrott, B. *J. Am. Chem. Soc.* **1992**, *114*, 2652−2656.
(30) Lohrmann, R.; Orgel, L. E. *Tetrahedron* **1978**, *34*, 853−855.
(31) Kanavarioti, A.; Lee, L. F.; Gangopadhyay, S. *Origins Life Evol. Biospheres* **1999**, *29*, 473−487.
(32) Verlander, M. S.; Lohrmann, R.; Orgel, L. E. *J. Mol. Evol.* **1973**, *2*, 303−316.
(33) Ferris, J. P.; Hill, A. R.; Liu, R.; Orgel, L. E. *Nature* **1996**, *381*, 59−61.
(34) Monnard, P.-A.; Kanavarioti, A.; Deamer, D. W. *J. Am. Chem. Soc.* **2003**, *125*, 13734−13740.
(35) Mariani, A.; Sutherland, J. D. *Angew. Chem., Int. Ed.* **2017**, *56*, 6563−6566.
(36) Usher, D. A.; McHale, A. H. *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73*, 1149−1153.
(37) Bowler, F. R.; Chan, C. K. W.; Duffy, C. D.; Gerland, B. B. B. B.; Islam, S.; Powner, M. W.; Sutherland, J. D.; Xu, J. *Nat. Chem.* **2013**, *5*, 383−389.
(38) Sutherland, J. D. *Nat. Rev. Chem.* **2017**, *1*, 12.

DOI: 10.1021/jacs.7b08784
J. Am. Chem. Soc. 2017, 139, 16741−16747